

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:47:02 ; Search time 210.42 Seconds  
(without alignments)  
108.227 Million cell updates/sec

Title: US-09-331-631A-3  
Perfect score: 3532  
Sequence: 1 MAINTSNLCSLFLSLFL.....SSRSRKQOQPLVSLDPVGF 666

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID        | Description            |
|------------|--------|-------------|--------|-----------|------------------------|
| 1          | 3532   | 100.0       | 666    | 19 W62829 | Macadamia integrifolia |
| 2          | 3412   | 96.6        | 666    | 19 W62828 | Macadamia integrifolia |
| 3          | 3215   | 91.0        | 625    | 19 W62830 | Macadamia integrifolia |
| 4          | 1145   | 32.4        | 566    | 13 R20181 | Sequence encoded b     |
| 5          | 1105.5 | 31.3        | 590    | 19 W62832 | Gossypium hirsutum     |
| 6          | 1036   | 29.3        | 525    | 19 W62831 | Theobroma cacao an     |
| 7          | 895    | 25.3        | 593    | 19 W62835 | Zea mays antimicro     |
| 8          | 875    | 24.8        | 637    | 19 W62837 | Hordeum vulgare an     |
| 9          | 873.5  | 24.7        | 614    | 18 W2149  | Peanut allergen Ar     |
| 10         | 873.5  | 24.7        | 614    | 19 W62834 | Archis hypogaea a      |
| 11         | 865.5  | 24.5        | 626    | 20 Y15244 | Peanut allergen A      |
| 12         | 865.5  | 24.5        | 626    | 20 Y25657 | Peanut allergen 11     |

|    |       |      |      |           |                       |
|----|-------|------|------|-----------|-----------------------|
| 13 | 859.5 | 24.3 | 626  | 18 W2150  | Peanut allergen Ar    |
| 14 | 855.5 | 24.2 | 605  | 19 W62838 | Glycine max antim     |
| 15 | 839.5 | 23.8 | 605  | 20 Y40999 | Soybean beta-congl    |
| 16 | 798   | 22.6 | 524  | 20 W90339 | G. max SBP1 protei    |
| 17 | 797.5 | 22.6 | 489  | 20 W90341 | G. max SBP2 protei    |
| 18 | 739   | 20.9 | 444  | 20 W90340 | G. max truncated S    |
| 19 | 727.5 | 20.6 | 415  | 20 Y40913 | Ara h 1 allergen p    |
| 20 | 689.5 | 19.5 | 409  | 20 W90342 | G. max truncated S    |
| 21 | 530.5 | 15.0 | 371  | 20 Y40914 | T. gondii immunoge    |
| 22 | 237   | 6.7  | 611  | 20 Y29039 | A human trichothal    |
| 23 | 226   | 6.4  | 1898 | 20 Y30795 | Leucocytozoan prot    |
| 24 | 218.5 | 6.2  | 1162 | 21 Y5500  | Human thyrotropin     |
| 25 | 187.5 | 5.3  | 562  | 16 R70491 | Human follicle sti    |
| 26 | 184   | 5.2  | 412  | 17 W03626 | Modified oat gliadi   |
| 27 | 183.5 | 5.2  | 1239 | 20 Y55931 | Rice storage prote    |
| 28 | 183   | 5.2  | 360  | 17 W03627 | Peanut allergen, A    |
| 29 | 173   | 4.9  | 472  | 15 R47127 | Ara h 3 allergen s    |
| 30 | 171   | 4.8  | 499  | 9 P82755  | P. falciparum LSA-R   |
| 31 | 170   | 4.8  | 510  | 20 Y15246 | Human CDC28-#3 RNA    |
| 32 | 169   | 4.8  | 512  | 20 Y40912 | Mouse STE20-relate    |
| 33 | 166   | 4.7  | 493  | 13 R26944 | Amino acid sequenc    |
| 34 | 164.5 | 4.7  | 316  | 13 R26941 | Human protein regulat |
| 35 | 163.5 | 4.6  | 1233 | 20 Y55954 | Oat globin A2B. A     |
| 36 | 163   | 4.6  | 1135 | 21 Y68784 | Full length human     |
| 37 | 161   | 4.6  | 1041 | 20 W30613 | Human ZC3 protein.    |
| 38 | 160   | 4.5  | 1299 | 21 Y58633 | Human GEX2 protein    |
| 39 | 159.5 | 4.5  | 515  | 15 R47128 | Biorhythmic marker p  |
| 40 | 157   | 4.4  | 968  | 20 Y55966 | Human ZC2 protein.    |
| 41 | 154.5 | 4.4  | 1326 | 20 Y55933 | LexA/NUMA fusion p    |
| 42 | 147   | 4.2  | 912  | 20 Y55939 |                       |
| 43 | 146.5 | 4.1  | 361  | 20 Y22206 |                       |
| 44 | 146.5 | 4.1  | 1297 | 20 Y55932 |                       |
| 45 | 146.5 | 4.1  | 2192 | 18 W21732 |                       |

#### ALIGNMENTS

|        |  |
|--------|--|
| RESULT | 1  |
| ID     | W62829 standard: Protein: 666 AA.                      |
| XX     |  |
| AC     | W62829;  |
| XX     |  |
| DT     | 27-OCT-1998 (first entry)                              |
| XX     |  |
| DE     | Macadamia integrifolia antimicrobial protein.          |
| XX     |  |
| KW     | antimicrobial protein; infestation; control.           |
| XX     |  |
| OS     | Macadamia integrifolia.                                |
| XX     |  |
| FH     | Key  |
| FT     | Peptide  |
| FT     | Protein  |
| FT     | Location/Qualifiers                                    |
| FT     | /note="signal peptide"                                 |
| FT     | 29..666  |
| FT     | /note="mature protein"                                 |
| XX     |  |
| PN     | W09827805-A1.  |
| PD     | 02-JUL-1998.   |
| XX     |  |
| PF     | 22-DEC-1997: 97WO-AU00874.                             |
| XX     |  |
| PR     | 20-DEC-1996: 96AU-0004275.                             |
| XX     |  |
| PA     | (RETR-) COOP RBS CENT TROPICAL PLANT PATHOLOGY.        |
| PI     | Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP: |
| DR     | WPI: 1998-377279/32.                                   |
| DR     | N-PSDB: V42311.  |
| XX     |  |

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
XX useful for controlling microbial infestations of plants or mammals  
XX  
PS Claim 1; Page 39-41; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 666 AA;

Query Match 100.0%; Score 3532; DB 19; Length 666;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 666; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAINTSNLCSLFLSLFLSTVSLAESFDRQYEECKRQCMOLETSQMRVCSQCD 60  
DB 1 maintsnlcsllflslstvtslaesefdrqyeckrqcmqltsqmrvcvsqd 60  
QY 61 KRFEEDIDMSKYNDODPQTDCCQCORRCROESGPRQOYQCRCKEICEEEYNNQR 120  
DB 61 krfeedidwskynqddpqtcdqccqrccrqesgprqqycqrckelceeeeynqr 120  
QY 121 DPQOQTECCOECORHETPRHMQTCQRCERYEKKRKQKRYEQOFEDEKTEERM 180  
DB 121 dpqdyeqecqcrqheteprrhmctqrceryekrkqkryeqqredeekyeerm 180  
QY 181 KEEDNKRDPQREYEDCRRCEQOEPQOYQCORCEQORHGRGDLINPORGSGRY 240  
DB 181 keednkrdpqreyedcrrceqeqprqyqqrccqrgyrggdlinpqr9sgry 240  
QY 241 EECEKQSDNPPYFDEERSLSTRTEEGHISVLENFYGSKLRLAKNRYLLEANPNA 300  
DB 241 eegeeksdnpyfderslsttrteeeghsvlenfygskllralknyllleanpna 300  
QY 301 FVLPTHLDADAILLVYGGRCALKMIHRDNRESYNLECGDVIRIPAGTTFYLINRNNERL 360  
DB 301 fvlpthldadaillvgyrgalkmihndresynlecgdviripagtlfyllnrnnerl 360  
QY 361 HIAKFLQTTSPQYKEFPFAGGONPEPYLSTFSKELFALMTQAERLGVLGQOOREGV 420  
DB 361 hiakfqtstspqykeffpagggnpepylstskelleaalntgaerlrgvlgqqrey 420  
QY 421 IISASQEQIRELTRDSESRHWHIRRGESSRGPYLFNKRPLYSNKYGOAYEVKPEDYR 480  
DB 421 iiasqeqireltrdsestrwhirrgessrgpylnfnkrplysnkygqayevkpedyr 480  
QY 481 QLODMYVSFIANITOGSMNGPFNTBSTKVVVYVVSAGADVEMACPHLSGRHGRGGR 540  
DB 481 qlqdmvsvfianltqgsnmngpfntbstkvvvvvasgaadvemacphlsgrhgrggr 540  
QY 541 HEEEDVHYEQVAKRLSKREAIIVPVGHVVFSSGMENTLFAFGNAGNNHENFLAGR 600  
DB 541 heeedvhyeqvakrlskrealivpvghvvfssgentllfafgnagnnhenflagr 600  
QY 601 ERNVLOQIEPOAMELFAAPRKEVEELFNQSDSIFFPGRQHQOOSRSTKQOQPLVSI 660  
DB 601 ervnlqiepqamelafaprkeveelfnsqdesifffpgrqhqqasrstkqqlvsi 660  
QY 661 LDFVGF 666  
DB 661 ldfvgf 666

RESULT 2  
ID W62828 standard; Protein; 666 AA.  
XX  
AC W62828;  
XX  
DT 27-OCT-1998 (first entry)  
XX

DE Macadamia integrifolia antimicrobial protein.  
XX  
KW antimicrobial protein; infestation; control.  
XX  
OS Macadamia integrifolia.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..28  
FT Protein /note= "signal peptide"  
FT 29..666  
FT /note= "mature protein"

PN W09827805-A1.  
PD 02-JUL-1998.  
PF 22-DEC-1997; 97WO-AU00874.  
PR 20-DEC-1996; 96AU-0004275.  
PA (KENTR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP.  
DR WPI; 1998-377279/32.  
DR N-PSDB; V42310.  
XX  
XX Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
XX Claim 1; Page 34-36; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 666 AA;

Query Match 96.6%; Score 3412; DB 19; Length 666;  
Best Local Similarity 96.1%; Pred. No. 1.5e-298;  
Matches 640; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 1 MAINTSNLCSLFLSLFLSTVSLAESFDRQYEECKRQCMOLETSQMRVCSQCD 60  
DB 1 maintsnlcsllflslstvtslaesefdrqyeckrqcmqltsqmrvcvsqd 60  
QY 61 KRFEEDIDMSKYNDODPQTDCCQCORRCROESGPRQOYQCRCKEICEEEYNNQR 120  
DB 61 krfeedidwskynqddpqtcdqccqrccrqesgprqqycqrckelceeeeynqr 120  
QY 121 DPQOQTECCOECORHETPRHMQTCQRCERYEKKRKQKRYEQOFEDEKTEERM 180  
DB 121 dpqdyeqecqcrqheteprrhmctqrceryekrkqkryeqqredeekyeerm 180  
QY 181 KEEDNKRDPQREYEDCRRCEQOEPQOYQCORCEQORHGRGDLINPORGSGRY 240  
DB 181 keednkrdpqreyedcrrceqeqprqyqqrccqrgyrggdlinpqr9sgry 240  
QY 241 EECEKQSDNPPYFDEERSLSTRTEEGHISVLENFYGSKLRLAKNRYLLEANPNA 300  
DB 241 eegeeksdnpyfderslsttrteeeghsvlenfygskllralknyllleanpna 300  
QY 301 FVLPTHLDADAILLVYGGRCALKMIHRDNRESYNLECGDVIRIPAGTTFYLINRNNERL 360  
DB 301 fvlpthldadaillvgyrgalkmihndresynlecgdviripagtlfyllnrnnerl 360  
QY 361 HIAKFLQTTSPQYKEFPFAGGONPEPYLSTFSKELFALMTQAERLGVLGQOOREGV 420  
DB 361 hiakfqtstspqykeffpagggnpepylstskelleaalntgaerlrgvlgqqrey 420  
QY 421 IISASQEQIRELTRDSESRHWHIRRGESSRGPYLFNKRPLYSNKYGOAYEVKPEDYR 480  
DB 421 iiasqeqireltrdsestrwhirrgessrgpylnfnkrplysnkygqayevkpedyr 480

Db 421 lirasgeqireltdseshwhlrrgessrgpynlfnkrlpysnkysgagvewkpedyr 480  
QY 481 QLODMADVFLANITOGSMGPFNTRSTKVVVVASGADVEMACPHLSGRHGRGGR 540  
Db 481 qlqmdlsvflantqgsmgpfntstrtkvrvvasgeadvemacphlsgrhgrggr 540  
QY 541 HEEEDVHYEOKARLSKREAIIVPVGHVYVSSGNENLLFAFGINAOHNENFLAGR 600  
Db 541 heeedvhyeqvarliskrealivlaghpvfvassgneillfajinaqnhenflagr 600  
QY 601 ERNVLQOIEPQAMELAPARKEVEELFNSODESIFPPGPHOHOQSSRSRKKOQPLVSI 660  
Db 601 ernvlqiepqamelaapakeveesfnsqgsilffpgrqbqgsprstkqqplvsi 660  
QY 661 LDFVGF 666  
Db 661 ldfvgf 666  
RESULT 3  
W62830  
ID W62830 standard; Protein; 625 AA.  
XX  
AC W62830;  
XX  
DT 27-OCT-1998 (first entry)  
XX  
DE Macadamia integrifolia antimicrobial protein.  
XX  
KM antimicrobial protein; infestation; control.  
XX  
OS Macadamia integrifolia.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..28 /note="signal peptide"  
FT Protein 29..666 /note="mature protein"  
XX  
PN WO9827805-A1.  
XX  
PD 02-JUL-1998.  
XX  
PF 22-DEC-1997; 97WO-AU00874.  
XX  
PR 20-DEC-1996; 96AU-0004275.  
XX  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
XX  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
XX  
DR WPI: 1998-377279/32.  
XX  
DR N-PSDB; V42316.  
XX  
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
XX  
PS useful for controlling microbial infestations of plants or mammals  
XX  
PS Claim 1; Page 43-45; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
XX  
CC be used to control microbial infestations in plants and mammalian  
XX  
CC animals.  
XX  
SQ Sequence 625 AA;  
Query Match 91.0%; Score 3215; DB 19; Length 625;  
Best Local Similarity 96.6%; Pred. No. 7.4e-281;  
Matches 604; Conservative 7; Mismatches 14; Indels 0; Gaps 0;  
QY 42 OCMDLETSGOMRRCVSGCDRPFEDIDMSKYDNODPDQCCOCORRCRQESGPRQOY 101  
Db 1 qcmdletsqgmrrcvsgcdkrfteedlmskysdndpdqccqrcrqesdprqy 60

QY 102 CORCKEICEEEENRORDPOOYEQOQERQNHETEPHMOQOQRCERREKERKQ 161  
Db 61 cqrckceiceeeeynqrpqgqyqeqckrqrcreteprhmqicqrcerreykerkq 120  
QY 162 OKRIEEOQREDEEYERKAEEDNKRDPQOREYEDCRRCEQOOPROOYCCORRCRQOR 221  
Db 121 qkryeeqgredeekyeermkgdnkrdpqgrevedcrrheqgeprtygcqrcrqeqqr 180  
QY 222 OHGRCGLINPORGSGREGEKEKSDNPYYRDERLSRFRTEEGHISVLNPNFRGSK 281  
Db 181 qhgrgqlmnpqrgsgryeegekgsdnpyyiderslsrfrteeahisvlenfygrsk 240  
QY 282 LLRALKNYRLVLEANEPAFVLPTHLDADAILLVTGGRGALKMIHRDNRESYNLECGDVI 341  
Db 241 llralknryrlvleapnafvlpthldadallivtgrgalkmihndresynlecgdvi 300  
QY 342 RIPAGTTFYILNRDNNEHLIAKFLQTISTRPGYKEFFPAGGQNPBYLSTFSKEILEAA 401  
Db 301 ripagtlfyllnrddnerlhiakflqlstspgykeffpaggqnpelystfskeillea 360  
QY 402 LMQAERLRGLGOORGVLIISASOEIRELTRDSDSRWHIRRGESSRGPNLFNKR 461  
Db 361 lmqaelrgvlqgrqegvliisaseqireltdseshwhlrrgessrgpynlfnkr 420  
QY 462 PLYSNKYGQAYEYKPEDYRLOQMDVSVFLANITOGSMGPFNTRSTKVVVVASGADV 521  
Db 421 plysnkygqayekpedyrlqmdvsvflantqgsmgpfntstrtkvrvvasgeadv 480  
QY 522 EMACPHLSGRHGRGGRKHEEEDVHYEOKARLSKREAIIVPVGHVYVSSGNENLL 581  
Db 481 emacphlsgrhgrggrkheeeevhyeqvarliskrealivlaghpvfvassgneill 540  
QY 582 LPAFGINAONNHNENFLAGRENRVLOQIEPQAMELAPARKEVEELFNSODESIFPPGR 641  
Db 541 lfafginaqnhenflagrenvlqiepqamelaapakeveesfnsqgsilffpgr 600  
QY 642 OHQOQSSKSTKQOQPLVSIIDFVGF 666  
Db 601 qhgqgsprstkqqplvsiidfygf 625  
RESULT 4  
R20181  
ID R20181 standard; Protein; 566 AA.  
XX  
AC R20181;  
XX  
DT 16-APR-1992 (first entry)  
XX  
DE Sequence encoded by 67 kD T. cacao protein cDNA.  
XX  
DE Cocoa; flavour; vicillin; seed storage protein.  
XX  
OS Theobroma cacao.  
XX  
OS WO9119801-A.  
XX  
PN WO9119801-A.  
XX  
PD 26-DEC-1991.  
XX  
PF 07-JUN-1991; 91WO-GB00914.  
XX  
PR 11-JUN-1990; 90GB-0013016.  
XX  
PA (MRSC ) MARS UK LTD.  
XX  
PI Spencer ME, Hodge R, Deakin EA, Ashton S;  
XX  
DR WPI: 1992-024418/03.  
XX  
DR N-PSDB; Q20377.  
XX  
PT Recombinant cocoa proteins - are responsible for flavour in cocoa  
beans and produced in large quantities using yeast and bacterial





Db 260 rllgrtngdqgllivraleeetrelrthascegghpmlppfge-srpgyslldqrpisla 318  
 QY 466 NKYGQAYEVKPEDYROLQDMDVSVFIANTTQSGMCGPFNTSTKVVVVASGADVEMAC 525  
 Db 319 nqgylqyeadaarsfhdaehdvsfanltagsmsapllftrtfrklayvpngkygaeivc 378  
 QY 526 PHLSGRHG---GRGGRKRHEEEDVH-----YEQVKARLSKREAIIVPVCHPV 571  
 Db 379 phrgsggeseerdkgyrtseeeeseegeaaggqyhlrlarlsptafvvpaghpfv 438  
 QY 572 FVSSGEMNLLLFAGINAOONNHENFLAGRNNLQOIEPQAMELAPAPRKEVELEFNSQ 631  
 Db 439 avastsnlqivcfevhadnkvflagad-nvlgkldrvakalsfaskaeewdevlgr 497  
 QY 632 DESIFPPGPRQ---HQOQSSRSTKQOQ 655  
 Db 498 rekgyflpgrpeesgysgheeregeere 524

RESULT 8  
 W62837  
 ID W62837 standard; Protein: 637 AA.  
 XX  
 AC W62837;  
 XX  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Hordeum vulgare antimicrobial protein.  
 XX  
 KM antimicrobial protein; infestation; control.  
 XX  
 OS Hordeum vulgare.  
 XX  
 PN W09827805-A1.  
 XX  
 PD 02-JUL-1998.  
 XX  
 PF 22-DEC-1997; 97WO-AU00874.  
 XX  
 PR 20-DEC-1996; 96AU-0004275.  
 XX  
 PA (REFR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
 XX  
 PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
 XX  
 DR WPI: 1998-377279/32.  
 XX  
 PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
 XX useful for controlling microbial infestations of plants or mammals  
 XX  
 PS Claim 1; Page 60-62; 96pp; English.  
 XX  
 CC The sequence is that of an antimicrobial protein which can  
 XX be used to control microbial infestations in plants and mammalian  
 XX animals.  
 CC  
 SQ Sequence 637 AA;

Query Match 24.8%; Score 875; DB 19; Length 637;  
 Best Local Similarity 35.6%; Pred. No. 3.5e-70;  
 Matches 227; Conservative 99; Mismatches 206; Indels 106; Gaps 22;

QY 70 SKYDNDPQ--TDCCQCRRCROESGPRQOYCORCKEIEEEEEYVNRDPQOYE 127  
 Db 27 ashdeddtrrgshlsqcvgrcgrer--pr---ysnarvcge-----rddqgh- 71  
 QY 128 QOERCQRHETEFRAHQTCQQRCEERYEKEKRRQQRKYEEQOQDEE-----KYERMK 182  
 Db 72 -----grhgeegqgrgvgwhgegeereehgrgrhgegeereehgrgrgwhgegere 125  
 QY 183 EDNKRDPQOREIEDCRRRCRQOEPRQOYOCQRCRQOQROHGNG---GDLLNPQRGSG 238

Db 126 eegrgy-----hgrhgege-----reeerggrgrhgegereegrg 165  
 QY 239 RVEEGE--EKQSDN--PYVEDERSLSTRFTEGCHISLVLENFNGRSLLBALKNYRLVLT 294  
 Db 166 rtgeegdeeqdsrrpyvgrpsfrtrllsqdngfralrpfqvsrllrglrdyvaim 225  
 QY 295 EANDPNAFVLPTHLADAILLVTSGRGALKMTIHRDNRESYNLECGDVRIRPACTTFYILNR 354  
 Db 226 evnprafivpgfdadgdyvvaegevlvtlengekrsylvkegdvylvapgslmhlant 285  
 QY 355 DNNERLIAFLQITISPGCYKEFPAGGONPEPPIYSTFSKELLEALNTQARLKGVLG 414  
 Db 286 dgrtklviakllhltsvpgkf-qfl-----svkpllaslskrylraafktslderlfn 339  
 QY 415 QO-----REGVLIISAQEOIRELTRDSE---SRMHIRGESSSGPYNLFNKRPVLS 465  
 Db 340 grgqgektsvslvrasseqlrelrtaaeaggghwppftrgdsrtdlnllegtrpka 399  
 QY 466 NKYGQAYEVKPEDYROLQDMDVSVFIANTTQSGMCGPFNTSTKVVVVASGADVEMAC 525  
 Db 400 nrgllyeadarsfhdaeqdvrvavantlpgsmtlapylnltsfklavlllegevqlvc 459  
 QY 526 PHL-----SGRHGGRGGRKRHEEED-----VHYQVKARLSKRE 560  
 Db 460 phlgreseserehngk--grrreeeddqrrrrgseseseeseeeqqdyelvararsgs 517  
 QY 561 AIIVPVGHPVVFVSS--GNENLLFAGINAOONNHENFLAGRNNVLOQIEPQAMELAF 618  
 Db 518 aivvppghvveissqsgsnlqvctelnaermerwlagr-nvlygklspsageltfg 576  
 QY 619 APRKEVELEFNSQDESI-FPPGPRQHQOQSSRSTKQOQ 655  
 Db 577 rpatevgevfragdqdegfvagp-----eqgsregeqg 610

RESULT 9  
 W22149  
 ID W22149 standard; Protein: 614 AA.  
 XX  
 AC W22149;  
 XX  
 DT 29-DEC-1997 (first entry)  
 XX  
 DE Peanut allergen Ara hi.  
 XX  
 KM Peanut; seed storage protein; allergen; allergy; hypersensitivity;  
 XX vaccine; anaphylactic shock; immunotherapy; therapy.  
 KM monoclonal antibody; ELISA; analysis; Ara hi.  
 XX  
 OS Arachis hypogaea strain Florunner.  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 1..22  
 FT Protein /label= Sig\_peptide  
 FT Modified-site 23..614  
 FT /label= Mat\_protein  
 FT 521..523  
 FT /note= "N-glycosylation site"  
 XX  
 PN W09724139-A1.  
 XX  
 PD 10-JUL-1997.  
 XX  
 PF 23-SEP-1996; 96WO-US15222.  
 XX  
 PR 04-MAR-1996; 96US-0610424.  
 PR 29-DEC-1995; 95US-0009455.  
 XX  
 PA (UYAR-) UNIV ARKANSAS.  
 XX  
 PI Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;  
 XX









|           |     |  |      |
|-----------|-----|--|------|
| OY        | 93  | ESGPFOOY---CORRKEICEEEEEENRORDDPOOQYEOCORRETPR-----H         | 142  |
| Db        | 26  | ksppqkktlmpcaqrcrlqscgcqppdlkqk-----acesctkleyprevcydbrgh    | 78   |
| OY        | 143 | MOTCOORC---ERR-----YEKEKKROOKRIEEOQREDEBEKEEPMKEDMKRDPQOR    | 1922 |
| Db        | 79  | tgltttrpppgeterrgrpdgyddtr--prrteagrtwpgaprrerered--wrgpr--  | 1333 |
| OY        | 193 | EYEDCRRCCEOEPRPOOYQOCORRCREOORHGRGDDLINFROGSGRGEEGEEKOSDNPY  | 252  |
| Db        | 134 | --edtrrrpsbq-pr-----krlpregege-----qewglpsbshvretstnmpf      | 175  |
| OY        | 253 | YFDEKSTRETRTEEGHISYLENFYGRSKLLRLAKNRLVLEENPAAEVLPTHLDAAI     | 312  |
| Db        | 176 | yfrsfrfstrfygnqgrttrvtrqtrfdgrsfqnlgnhrivgleakprotlrlphdadnl | 235  |
| OY        | 313 | LLVYGRGALAKMIHRDNRESYNELECGDVIRIAGTFTYLLINRDNNEERLHAKLQITSTP | 372  |
| Db        | 236 | lvlgqgqacvtevangnmrksnldegalltrpsfisyllnrdnqlvrvaklsmpvtqp   | 295  |
| OY        | 373 | GOYKEFPFAGQONDEPYLSTFSKEILLAAUNTOAERLGVL-----GOOR---         | 417  |
| Db        | 296 | qgfedfppasrrdqsylqgfsnrltlaaInaelneltvllleanggegeeqgrrrwt    | 355  |
| OY        | 418 | -----BCVITISASOEOIRLRTDSDSRKMHTRGCESSRG-----PYNLFMKRPLYSNK   | 467  |
| Db        | 356 | rsseeneqvllvkakehveeltkhaks-----vskksseegdltnpInlregepdlsm   | 410  |
| OY        | 468 | YGOAYEYKPEYR-OLODMDVSVTIANITOGSMKGPENRSTKRVVYVVASGEADVEMACP  | 526  |
| Db        | 411 | fgklfevkrpdkknpqlqdlmmltcveikegalmlpfnfskamyivvynkgtnlelvay  | 470  |
| OY        | 527 | HLSGHGRGRGGRKHEEEDV---HYEQVKARLSKREAIIVPYGHPVFEVSSGNENLL     | 582  |
| Db        | 471 | rkeqqrqrteedeedeeseegsnrevrlytarlkegdvflmpaahpvaInassehlh    | 529  |
| OY        | 583 | FAGFGMAQNNNEENFLAGERVLYLOQIEPOMELAFAPRKREVELFNSQDESTFFPGPRQ  | 642  |
| Db        | 530 | -gfIgaeneenmrlllaadkdhvldjqekxadlafpsgqeveklXkngkeshlfsaarpg | 588  |
| OY        | 643 | HQOQSSTSTKOOP  | 656  |
| Db        | 589 | sqsqspespekesp   | 602  |
| RESULT 14 |     |  |      |
| W62838    | ID  | W62838 standard; Protein; 605 AA.                            |      |
| XX        | AC  | W62838;  |      |
| XX        | DT  | 27-OCT-1998 (first entry)                                    |      |
| XX        | GE  | Glycine max antimicrobial protein.                           |      |
| XX        | KW  | antimicrobial protein; infestation; control.                 |      |
| XX        | OS  | Glycine max.   |      |
| XX        | PN  | W09827805-A1.  |      |
| XX        | PD  | 02-JUL-1998.   |      |
| XX        | PF  | 22-DEC-1997; 97MO-A000874.                                   |      |
| XX        | PR  | 20-DEC-1996; 96AU-0004275.                                   |      |
| XX        | PA  | (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.              |      |
| XX        | PI  | Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP.       |      |
| XX        | DR  | WPI; 1996-377279/32.   |      |

PT Novel antimicrobial protein from e.g. *Macademia integrifolia* -  
 PT useful for controlling microbial infestations of plants or mammals  
 XX  
 XX  
 PS Claim 1; Page 63-65; 96pp; English.  
 XX  
 XX  
 CC The sequence is that of an antimicrobial protein which can  
 CC be used to control microbial infestations in plants and mammalian  
 CC animals.  
 XX  
 XX  
 SQ Sequence 605 AA;

|                       |                   |                    |            |             |
|-----------------------|-------------------|--------------------|------------|-------------|
| Query Match           | 24.2%;            | Score 855.5;       | DB 19;     | Length 605; |
| Best Local Similarity | 31.4%;            | Pred. No. 1.8e-68; |            |             |
| Matches 194;          | Conservative 143; | Mismatches 211;    | Indels 69; | Gaps 15     |

```
QY      72 YDNDDDPQCDCQQCQRRCRQESGPRDQYCYQARC-----KEIEEEE---EYNRRDP 122
          |::|::||| |||||
Db     27 ywekenpkim--kcqscnsedcsyrnqa-charcnllkvekeceegelprrprpqhp 83
```

```

0y 123 qqvveccqecrnhfeprhmtcqcncerrveyekrkkooryeeqoredekeyeenke 182
    :: :| : : :|| : : : | |:|::| : :: : | :| : :
Db 84 erepqpgkexedeedeqprlpfprrpqreeehqreeqewprkeekrgkygseedd 143

```

```
QY      183 EDNNRDPQGRREV EDCRRRCEDQEPRDQVQCQRRCREDDQNGRGSLINPORGSSGVEE 242
        ||::||:| |::||::||:|
Db     144 edeeqdergfprp---rphngkeerneeded--eeqges-----ee 181
```

```
Dy      243 GEEKQ-----SDNPYFEDERSLSTRFTTEEGHISVLENNFYGRSKLLRALKNVRVLVEAN 2970
        :   :     :    |       |   |   |   |   |   |   |   |   |   |   |   |
Db      182 sedselrrhkknkpfifgsnrftellfkngryirlvlgfrinqrpsq|gnldirylllefnsk 241
```

Db 242 pntlllpnhadadyllivlingtalislvnnddcsyrlqsgdalrvpsgtlyyvvnpdm 3011

QY 358 ERRLIAKFGTISTGQYKEEPAGCONPEPVYSTFSKELLEALUMQAERLGVL---- 413  
 | | : :: ||:: || : : || ||: ||||: :: | : ||  
 Db 302 enrlrlaivpnkpyrfesflstteaqsylqgfsrmlleasydlkfeelinkvlfsre 361

```

414 -----GGGRR--EGVLLIASQCEIRLRLDDSESRKWHIRKGGESSRGPNYFNENKRPYSN 466
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
db 362 eegqgqgeqlqgesviveiskqiralskrakssrkti---sseedkpfnlrlsrpdlpsyn 4177

```

```

40/ KICGAEVAFEDIRLODMDSVF IANTIGSMNGPFETKRSIKVAVASAEADVACAP 526
|:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 418 klqkfelfepknpqlrldlflslvmnegallphfmskalvllvinegdanielv-- 475

```

```

521  HUSGRHSG--KRGGRHHEEELEVNHELVNARKSLNREALIVPVGRHVVVFSSGNNELILFA 384
      | : :: || : : | ||:: | : | ||| : : || |
476  -----gkKegqeqqeqeplvrvkyraelsedlflvpagypvv--natsnlnffa 526

```

```

527 1g1naemqrnflaegsqdnv1sqipsgyqela1pysagavek1lknqresfvdagpkkk 586

```

```
07      043 0008NSINQ00FLVSL 001  
          :: :: | |||  
Db      587 eegnkgrk--gplsl 601
```

|        |                           |
|--------|---------------------------|
| RESULT | 15                        |
| Y40999 |                           |
| Y40999 | standard: protein: 605 AA |

|    |                           |
|----|---------------------------|
| AA |                           |
| AC | Y40999;                   |
| XX |                           |
| DT | 06-DEC-1999 (first entry) |

AA Soybean beta-conglycinin protein sequence.  
 XX  
 DE  
 WU  
 Deannut; a]l]ercon; Aya H 1; Tof; immuno]obulin F; onfrono; Aya h 3.

allergic reaction; soybean; beta-conglycinin.

```
XX Glycine max.
OS
XX
XX MO9945961-A1.
XX
XX 16-SEP-1999.
XX
XX 12-MAR-1999; 99WC-US05494.
XX
XX 12-MAR-1998; 98US-0077763.
XX
XX 11-MAR-1999; 99US-0077763.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX Burks W, Helm RM, Cockrell G, Bannon GA, Stanley JS, Shin DS;
XX Sampson H, Compadre CM, Huang SK, Maleki SJ, Kopper RA;
XX WPI; 1999-551218/46.
XX
XX Tertiary structure of peanut allergen Ara h 1 for protection of a host
XX animal from allergic reaction -
XX
XX PS Disclosure; Fig 33A-B; 193pp; English.
XX
XX The invention provides a tertiary structure for the peanut allergen
XX Ara h 1. The Ara h 1 allergen is found to contain 23 linear IgE-binding
XX epitopes. The invention also provides an isolated recombinant peanut
XX allergen designated Ara h 3 and a nucleotide molecule encoding the peanut
XX allergen Ara h 3. Molecules of the invention are used to protect a host
XX animal from allergic reaction, particularly using a modified allergen
XX which is less reactive with IgE. The invention may also be used to
XX ensure that the allergen is not introduced into genetically modified
XX food. The present sequence represents a soybean beta-conglycinin protein
XX sequence.
SQ Sequence 605 AA;

Query Match 23.8%; Score 839.5; DB 20; Length 605;
Best Local Similarity 30.9%; Pred. No. 5e-67;
Matches 192; Conservative 142; Mismatches 212; Indels 75; Gaps 15;

QY 69 WSKYNDODDPOTDCQOCRCRQOESGPROOYQYCRCKELCEEEERYN-----RQ 119
DB 28 weknpxhn-----kclqscnserdsyrnqa-charcnllkveckekkgelprprp 80
QY 120 RDPQOQYEOCOERCORHETPRHMQTCOCERREYKEKRRKQOQRYEQOREDEKYEER 179
DB 81 qhperpqyqpekeedeegprpripfpqprqeehneqreeg---ewprkeekigekg 136
QY 180 MKEEDNKRDPQOREYE-DCRRRCQDEPRROOYQOCRCRCQORQHGRGDLINPQRGSG 238
DB 137 seededeedeegqerqfprfpqhpkqeeeneeded--eeqres----- 179
QY 239 RYEGGEKQ-----SDNPYFDERSLSTRTEEGHISVLENYGRSKLLRALKNRVL 293
DB 180 --eesedseftrhknkprflfsgnrfelfknygrirvlfqrxpsqlnrdyfile 237
QY 294 LEAPNFAFVPTHLDAADAILVTGGRGALMKIHNRNRESYNLECGDVIRIPAGTFEYLIN 353
DB 238 fnsfpnclllpnhadayllivngtalislvmddrdsyrgdaltvpsgtlyyvn 297
QY 354 RDNNERLHAKFLQTLSTPOAYKEFPFAGQNPPEYLSFSKEILEALNTQAERLGVL 413
DB 298 pdnmenrltllaipvknkpgfseflsfrdqgsylqfsmllleaaydtkfeelinkvl 357
QY 414 -----GOOR--EGVITASQEOJRELJTRDQSESRKWHIRGGESSRGPPYNLENKRP 462
DB 358 fsreegqgqgeqrlqesviveisqelralskrakssrktl----ssedkpfnlrsrdp 413
QY 463 LYSKKVGQAYEAVPEDRQLODMQVAFININOGSMGPFETRTSKVYVVASGEADVE 522
DB 414 lysnkigkffeipknpqrlididifislvdmegalllphfnskaivlivineganle 473
```

```
QY 523 MACPHLSGRHGC--RGGKRRHEEEDVHYEQVAKRLSKREALIVPVGHPVVFSSGNEL 580
DB 474 lv-----glkeqgqeqgeqplevkvkyraelsqdlfviipagypvvv--natsnl 522
QY 581 LLEAFGINAONNHENFLAGKERNVLAQIEPQAMELAFADPRKEVEELFNSQDSIFPPGP 640
DB 523 nffaiginaenqgnflagsqdhnvisqipsqvgelafpsaqavkklknqrksyfvdaq 582
QY 641 ROHQOQSSRSTKQOQPLVSTL 661
DB 583 pkkeegnkgrk--gpissll 601
```

Search completed: March 1, 2001, 15:47:07  
Job time: 232 sec

